

Pls Ex. 9  
(JCCX76 - Ronald G. Ehrenberg &  
Robert S. Smith, Modern Labor  
Economics: Theory and Public Policy  
(9th ed., 2006))

Part 1

# MODERN LABOR ECONOMICS

*Theory and Public Policy*

NINTH EDITION

**RONALD G. EHRENBURG**

*School of Industrial and Labor Relations  
Cornell University*

**ROBERT S. SMITH**

*School of Industrial and Labor Relations  
Cornell University*



Boston San Francisco New York  
London Toronto Sydney Tokyo Singapore Madrid  
Mexico City Munich Paris Cape Town Hong Kong Montreal

General Library System  
University of Wisconsin - Madison  
3 State Street  
Madison, WI 53706-1494  
U.S.A.

Vice President and Editorial Director: Daryl Fox  
Editor in Chief: Denise Clinton  
Acquisitions Editor: Adrienne D'Ambrosio  
Editorial Assistant: Ashley Booth  
Senior Production Supervisor: Meredith Gertz  
Supplements Editor: Kirsten Dickerson  
Executive Marketing Manager: Stephen Frail  
Design Manager: Regina Hagen Kolenda  
Media Producer: Bridget Page  
Composition, Illustration, and Packaging Services: Electronic Publishing Services Inc.,  
NYC  
Cover Designer: Leslie Haimes  
Senior Print Buyer: Hugh Crawford

Library of Congress Cataloging-in-Publication Data

Ehrenberg, Ronald G.

Modern labor economics : theory and public policy / Ronald G. Ehrenberg, Robert S. Smith.—  
9th ed.

p. cm.

Includes index.

ISBN 0-321-30503-5

1. Labor economics. 2. Labor policy. 3. Personnel management. I. Smith, Robert Stewart. II.

Title.

HD4901.E34 2005

331—dc22

2004028156

Copyright © 2006 by Pearson Education, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval  
system, or transmitted, in any form or by any means, electronic, mechanical, photocopy-  
ing, recording, or otherwise, without the prior written permission of the publisher.  
Printed in the United States of America.

12345678910—CRW—0908070605



## The Demand for Labor

The demand for labor is a derived demand, in that workers are hired for the contribution they can make toward producing some good or service for sale. However, the wages workers receive, the employee benefits they qualify for, and even their working conditions are all influenced, to one degree or another, by the government. There are minimum wage laws, pension regulations, restrictions on firing workers, safety requirements, immigration controls, and government-provided pension and unemployment benefits that are financed through employer payroll taxes. All these requirements and regulations have one thing in common: they increase employers' costs of hiring workers.

We explained in chapter 2 that both the scale and the substitution effects accompanying a wage change suggest that the demand curve for labor is a downward-sloping function of the wage rate. If this rather simple proposition is true, then policies that mandate increases in the costs of employing workers will have the undesirable side effect of reducing their employment

opportunities. If the reduction is large enough, lost job opportunities actually could undo any help provided to workers by the regulations. Understanding the characteristics of labor demand curves, then, is absolutely crucial to anyone interested in public policy. To a great extent, how one feels about many labor market regulatory programs is a function of one's beliefs about labor demand curves!

The current chapter will identify *assumptions* underlying the proposition that labor demand is a downward-sloping function of the wage rate. Chapter 4 will take the downward-sloping nature of labor demand curves as given, addressing instead why, in the face of a given wage increase, declines in demand might be large in some cases and barely perceptible in others.

## Profit Maximization

The fundamental assumption of labor demand theory is that firms—the employers of labor—seek to maximize profits (or, in the case of not-for-profit employers, some measure of services rendered, net of costs). In doing so, firms are assumed to continually ask, “Can we make changes that will improve profits?” Two things should be noted about this constant search for enhanced profits. First, a firm can make changes only in variables that are within its control. Because the price a firm can charge for its product and the prices it must pay for its inputs are largely determined by others (the “market”), profit-maximizing decisions by a firm mainly involve the question of *whether, and how, to increase or decrease output*.

Second, because the firm is assumed to constantly search for profit-improving possibilities, our theory must address the *small* (“marginal”) changes that must be made almost daily. Really major decisions of whether to open a new plant or introduce a new product line, for example, are relatively rare; once having made them, the employer must approach profit maximization incrementally through the trial-and-error process of small changes. We therefore need to understand the basis for these incremental decisions, paying particular attention to when an employer stops making changes in output levels or in its mix of inputs.

(With respect to the employment of inputs, it is important to recognize that analyzing marginal changes implies considering a small change in one input *while holding employment of other inputs constant*. Thus, when analyzing the effects of adjusting the labor input by one unit, for example, we will do so on the assumption that capital is held constant. Likewise, marginal changes in capital will be considered assuming the labor input is held constant.)

In incrementally deciding on its optimal level of output, the profit-maximizing firm will want to expand output by one unit if the added revenue from selling that unit is greater than the added cost of producing it. As long as the marginal revenue from an added unit of output exceeds its marginal cost, the firm will continue to expand output. Likewise, the firm will want to contract output whenever the marginal cost of production exceeds marginal revenue. Profits are maximized (and the firm stops making changes) when output is such that marginal revenue equals marginal cost.

A firm can expand or contract output, of course, only by altering its use of *inputs*. In the most general sense, we will assume that a firm produces its output by combining two types of inputs, or *factors of production*: *labor and capital*. Thus, the rules stated above for deciding whether to marginally increase or reduce output have important corollaries with respect to the employment of labor and capital:

- a. If the income generated by employing one more unit of an input exceeds the additional expense, then add a unit of that input;
- b. If the income generated by one more unit of input is less than the additional expense, reduce employment of that input;
- c. If the income generated by one more unit of input is equal to the additional expense, no further changes in that input are desirable.

Decision rules (a) through (c) state the profit-maximizing criterion in terms of *inputs* rather than output; as we will see, these rules are useful guides to deciding *how*—as well as *whether*—to marginally increase or decrease output. Let us define and examine the components of these decision rules more closely.

### Marginal Income from an Additional Unit of Input

Employing one more unit of either labor or capital generates additional income for the firm because of the added output that is produced and sold. Similarly, reducing the employment of labor or capital reduces a firm's income flow because the output available for sale is reduced. Thus, the marginal income associated with a unit of input is found by multiplying two quantities: the change in physical output produced (called the input's *marginal product*) and the *marginal revenue* generated per unit of physical output. We will therefore call the marginal income produced by a unit of input the input's *marginal revenue product*. For example, if the presence of a tennis star increases attendance at a tournament by 20,000 spectators, and the organizers net \$25 from each additional fan, the marginal income produced by this star is equal to her marginal product (20,000 fans) times the marginal revenue of \$25 per fan. Thus, her marginal revenue product equals \$500,000. (For an actual calculation of marginal revenue product in college football, see Example 3.1.)

**MARGINAL PRODUCT** Formally, we will define the *marginal product of labor*, or  $MP_L$ , as the change in physical output ( $\Delta Q$ ) produced by a change in the units of labor ( $\Delta L$ ), holding capital constant:<sup>1</sup>

$$MP_L = \Delta Q / \Delta L \quad (\text{holding capital constant}) \quad (3.1)$$

Likewise, the marginal product of capital ( $MP_K$ ) will be defined as the change in output associated with a one-unit change in the stock of capital ( $\Delta K$ ), holding labor constant:

$$MP_K = \Delta Q / \Delta K \quad (\text{holding labor constant}) \quad (3.2)$$

<sup>1</sup>The symbol  $\Delta$  (the uppercase Greek letter delta) is used to signify "a change in."

**EXAMPLE 3.1****The Marginal Revenue Product of College Football Stars**

Calculating a worker's marginal revenue product is often very complicated due to lack of data and the difficulty of making sure that everything else is being held constant and only *additions* to revenue are counted. Perhaps for this reason, economists have been attracted to the sports industry, which generates so many statistics on player productivity and team revenues.

Football is a big-time concern on many campuses, and some star athletes generate huge revenues for their colleges, even though they are not paid—except by receiving a free education. Robert Brown collected revenue statistics for 47 Division I-A college football programs for the 1988–1989 season—including revenues retained by the school from ticket sales, donations to the athletic department, and television and radio payments. (Unfortunately,

this leaves out some other potentially important revenue sources, such as parking and concessions at games and donations to the general fund.)

Next, he examined variation in revenues due to market size, strength of opponents, national ranking, and the number of players on the team who were so good that they were drafted into professional football (the National Football League [NFL]). Brown found that each additional player drafted into the NFL was worth about \$540,000 in extra revenue to his team. Over a four-year college career, a premium player could therefore generate over \$2 million in revenues for his team!

*Data from:* Robert W. Brown, "An Estimate of the Rent Generated by a Premium College Football Player," *Economic Inquiry* 31 (October 1993), 671–684.

**MARGINAL REVENUE** The definitions in (3.1) and (3.2) reflect the fact that a firm can expand or contract its output only by increasing or decreasing its use of either labor or capital. The marginal revenue (*MR*) that is generated by an extra unit of output depends on the characteristics of the product market in which that output is sold. If the firm operates in a purely competitive product market, and therefore has many competitors and no control over product price, the marginal revenue per unit of output sold is equal to product price (*P*). If the firm has a differentiated product, and thus has some degree of monopoly power in its product market, extra units of output can be sold only if product price is reduced (because the firm faces the *market* demand curve for its particular product); students will recall from introductory economics that in this case marginal revenue is less than price ( $MR < P$ ).<sup>2</sup>

**MARGINAL REVENUE PRODUCT** Combining the definitions presented in this subsection, the firm's marginal revenue product of labor, or  $MRP_L$ , can be represented as

$$MRP_L = MP_L \cdot MR \quad (\text{in the general case}) \quad (3.3a)$$

<sup>2</sup>A competitive firm can sell added units of output at the market price because it is so small relative to the entire market that its output does not affect price. A monopolist, however, is the supply side of the product market, so to sell extra output it must lower price. Because it must lower price on *all* units of output, and not just on the extra units to be sold, the marginal revenue associated with an additional unit is below price.

or as

$$MRP_L = MP_L \cdot P \quad (\text{if the product market is competitive}) \quad (3.3b)$$

Likewise, the firm's marginal revenue product of capital ( $MRP_K$ ) can be represented as  $MP_K \cdot MR$  in the general case, or as  $MP_K \cdot P$  if the product market is competitive.

### Marginal Expense of an Added Input

Changing the levels of labor or capital employed, of course, will add to or subtract from the firm's total costs. The marginal expense of labor ( $ME_L$ ) that is incurred by hiring more labor is affected by the nature of competition in the labor market. If the firm operates in a competitive labor market and has no control over the wages that must be paid (it is a "wage taker"), then the marginal expense of labor is simply the market wage. Put differently, firms in competitive labor markets have labor supply curves that are horizontal at the going wage (refer back to Figure 2.11); if they hire an additional hour of labor, their costs increase by an amount equal to the wage rate,  $W$ .

In this chapter, we will maintain the assumption that the labor market is competitive and that the labor supply curve to firms is therefore *horizontal* at the going wage. In chapter 5, we will relax this assumption and analyze how upward-sloping labor supply curves to individual employers alter the marginal expense of labor.

In the analysis that follows, the marginal expense of adding a unit of capital will be represented as  $C$ , which can be thought of as the expense of renting a unit of capital for one time period. The specific calculation of  $C$  need not concern us here, but clearly it depends on the purchase price of the capital asset, its expected useful life, the rate of interest on borrowed funds, and even special tax provisions regarding capital.

## The Short-Run Demand for Labor When Both Product and Labor Markets Are Competitive

The simplest way to understand how the profit-maximizing behavior of firms generates a labor demand curve is to analyze the firm's behavior over a period of time so short that the firm cannot vary its stock of capital. This period is what we will call the *short run*, and of course the time period involved will vary from firm to firm (an accounting service might be able to order and install a new computing system for the preparation of tax returns within three months, whereas it may take an oil refinery five years to install a new production process). What is simplifying about the short run is that, with capital fixed, a firm's choice of output level and its choice of employment level are two aspects of the very same decision. Put differently, in the short run the firm needs only to decide *whether* to alter its output level; *how* to increase or decrease output is not an issue, because only the employment of labor can be adjusted.

**TABLE 3.1**

The Marginal Product of Labor in a Hypothetical Car Dealership (capital held constant)

Number of Salespersons	Total Cars Sold	Marginal Product of Labor
0	0	
1	10	10
2	21	11
3	26	5
4	29	3

### A Critical Assumption: Declining $MP_L$

We defined the marginal product of labor ( $MP_L$ ) as the change in the (physical) output of a firm when it changes its employment of labor by one unit, holding capital constant. Since the firm can vary its employment of labor, we must consider how increasing or reducing labor will affect labor's marginal product. Consider Table 3.1, which illustrates a hypothetical car dealership with sales personnel who are all equally hardworking and persuasive. With no sales staff, the dealership is assumed to sell zero cars, but with one salesperson, it will sell 10 cars per month. Thus, the marginal product of the first salesperson hired is 10. If a second person is hired, total output is assumed to rise from 10 to 21, implying that the marginal product of a second salesperson is 11. If a third equally persuasive salesperson is hired, sales rise from 21 to 26 ( $MP_L = 5$ ), and if a fourth is hired, sales rise from 26 to 29 ( $MP_L = 3$ ).

Table 3.1 assumes that adding an extra salesperson increases output (cars sold) in each case. As long as output increases as labor is added, labor's marginal product is positive. In our example, however, the marginal product of labor increased at first (from 10 to 11), but then fell (to 5 and eventually to 3). Why?

The initial rise in marginal product occurs *not* because the second salesperson is better than the first; we ruled out this possibility by our assumption that the salespeople were equally capable. Rather, the rise could be the result of cooperation between the two in generating promotional ideas or helping each other out in some way. Eventually, however, as more salespeople are hired, the marginal product of labor must fall. A fixed building (remember that capital is held constant) can contain only so many cars and customers, and thus each additional increment of labor must eventually produce progressively smaller increments of output. This law of diminishing marginal returns is an empirical proposition that derives from the fact that as employment expands, each additional worker has a progressively smaller share of the capital stock to work with. For expository convenience, we shall assume that the marginal product of labor is always decreasing.<sup>3</sup>

<sup>3</sup>We lose nothing by this assumption, because we show later in this section that a firm will never be operated at a point where its marginal product of labor is increasing.

### From Profit Maximization to Labor Demand

From the profit-maximizing decision rules discussed earlier, it is clear that the firm should keep increasing its employment of labor as long as labor's marginal revenue product exceeds its marginal expense. Conversely, it should keep reducing its employment of labor as long as the expense saved is greater than the income lost. *Profits are maximized, then, only when employment is such that any further one-unit change in labor would have a marginal revenue product equal to marginal expense:*

$$MRP_L = ME_L \quad (3.4)$$

Under our current assumptions of competitive product and labor markets, we can symbolically represent the profit-maximizing level of labor input as that level at which

$$MP_L \cdot P = W \quad (3.5)$$

Clearly, equation (3.5) is stated in terms of some *monetary* unit (dollars, for example).

Alternatively, however, we can divide both sides of equation (3.5) by product price,  $P$ , and state the profit-maximizing condition for hiring labor in terms of *physical quantities*:

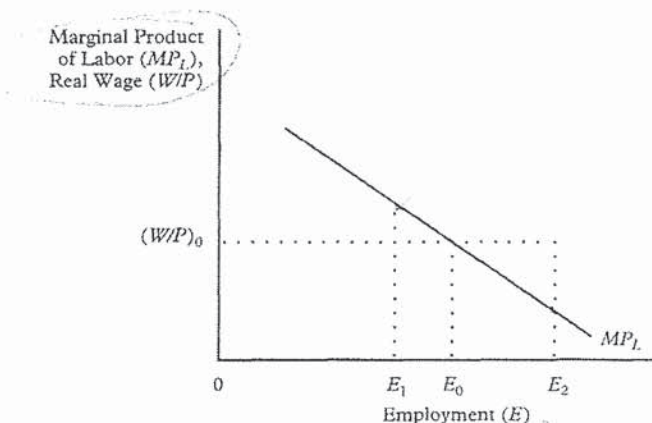
$$MP_L = W/P \quad (3.6)$$

We defined  $MP_L$  as the change in physical output associated with a one-unit change in labor, so it is obvious that the left-hand side of equation (3.6) is in physical quantities. To understand that the right-hand side is also in physical quantities, note that the numerator ( $W$ ) is the dollars per unit of labor and the denominator ( $P$ ) is the dollars per unit of output. Thus, the ratio  $W/P$  has the dimension of physical units. For example, if a woman is paid \$10 per hour and the output she produces sells for \$2 per unit, from the firm's viewpoint she is paid five units of output per hour ( $10 \div 2$ ). From the perspective of the firm, these five units represent her "real wage."

**LABOR DEMAND IN TERMS OF REAL WAGES** The demand for labor can be analyzed in terms of either *real* or *money* wages. Which version of demand analysis is used is a matter of convenience only. In this and the following subsection, we give examples of both.

Figure 3.1 shows a marginal product of labor schedule ( $MP_L$ ) for a representative firm. In this figure, the marginal product of labor is tabulated on the vertical axis and the number of units of labor employed on the horizontal axis. The negative slope of the schedule indicates that each additional unit of labor employed produces a progressively smaller (but still positive) increment in output. Because the real wage and the marginal product of labor are both measured in the same dimension (units of output), we can also plot the real wage on the vertical axis of Figure 3.1.

**FIGURE 3.1**  
Demand for Labor in the Short Run  
(Real Wage)



Given any real wage (by the market), the firm should thus employ labor to the point at which the marginal product of labor just equals the real wage (equation 3.6). In other words, the firm's demand for labor in the short run is equivalent to the downward-sloping segment of its marginal product of labor schedule.<sup>4</sup>

To see that this is true, pick any real wage—for example, the real wage denoted by  $(W/P)_0$  in Figure 3.1. We have asserted that the firm's demand for labor is equal to its marginal product of labor schedule and consequently that the firm would employ  $E_0$  employees. Now suppose that a firm initially employed  $E_2$  workers as indicated in Figure 3.1, where  $E_2$  is any employment level greater than  $E_0$ . At the employment level  $E_2$ , the marginal product of labor is less than the real wage rate; the marginal real cost of the last unit of labor hired is therefore greater than its marginal product. As a result, profit could be increased by reducing the level of employment. Similarly, suppose instead that a firm initially employed  $E_1$  employees, where  $E_1$  is any employment level less than  $E_0$ . Given the specified real wage  $(W/P)_0$ , the marginal product of labor is greater than the real wage rate at  $E_1$ —and consequently the marginal additions to output of an extra unit of labor exceed its marginal real cost. As a result, a firm could increase its profit level by expanding its level of employment.

Hence, to maximize profits, given any real wage rate, a firm should stop employing labor at the point at which any additional labor would cost more than it would produce. This profit-maximization rule implies two things. First, the firm should employ labor up to the point at which its real wage equals the marginal product of labor—but not beyond that point.

Second, its profit-maximizing level of employment lies in the range where its marginal product of labor is declining. If  $W/P = MP_L$  but  $MP_L$  is increasing, then

<sup>4</sup>We should add here, "provided that the firm's revenue exceeds its labor costs." Above some real wage level, this may fail to occur, and the firm will go out of business (employment will drop to zero).

adding another unit of labor will create a situation in which marginal product *exceeds*  $W/P$ . As long as adding labor causes  $MP_L$  to exceed  $W/P$ , the profit-maximizing firm will continue to hire labor. It will stop hiring only when an extra unit of labor would reduce  $MP_L$  below  $W/P$ , which will happen only when  $MP_L$  is declining. Thus, the only employment levels that could possibly be consistent with profit maximization are those in the range where  $MP_L$  is decreasing.

**LABOR DEMAND IN TERMS OF MONEY WAGES** In some circumstances, labor demand curves are more readily conceptualized as downward-sloping functions of *money* wages. To make the analysis as concrete as possible, in this subsection we analyze the demand for department store detectives.

At a business conference one day, a department store executive boasted that his store had reduced theft to 1 percent of total sales. A colleague shook her head and said, "I think that's too low. I figure it should be about 2 percent of sales." How can more shoplifting be better than less? The answer is based on the fact that reducing theft is costly in itself. A profit-maximizing firm will not want to take steps to reduce shoplifting if the added costs it must bear in so doing exceed the value of the savings such steps will generate.

Table 3.2 shows a hypothetical marginal revenue product of labor ( $MRP_L$ ) schedule for department store detectives. Hiring one detective would, in this example, save \$50 worth of thefts per hour. Two detectives could save \$90 worth of thefts each hour, or \$40 more than hiring just one. The  $MRP_L$  of hiring a second detective is thus \$40. A third detective would add \$20 more to thefts prevented each hour.

The  $MRP_L$  does *not* decline from \$40 to \$20 because the added detectives are incompetent; in fact, we shall assume that all are equally alert and well trained.  $MRP_L$  declines, in part, because surveillance equipment (capital) is fixed; with each added detective, there is less equipment per person. However, the  $MRP_L$  also declines because it becomes progressively harder to generate savings. With just a few detectives, the only thieves caught will be the more-obvious, less-experienced

**TABLE 3.2**

Hypothetical Schedule of Marginal Revenue Productivity of Labor for Store Detectives

Number of Detectives on Duty during Each Hour Store Is Open	Total Value of Thefts Prevented per Hour	Marginal Value of Thefts Prevented per Hour ( $MRP_L$ )
0	\$ 0	\$—
1	50	50
2	90	40
3	110	20
4	115	5
5	117	2

shoplifters. As more detectives are hired, it becomes possible to prevent theft by the more-expert shoplifters, but they are harder to detect and fewer in number. Thus,  $MRP_L$  falls because theft prevention becomes more difficult once all those who are easy to catch are apprehended.

To draw the demand curve for labor, we need to determine how many detectives the store will want to employ at a given wage. For example, at a wage of \$50 per hour, how many detectives will the store want? Using the  $MRP_L = W$  criterion (equation 3.5), it is easy to see that the answer is "one." At \$40 per hour, the store would want to hire two, and at \$20 per hour, the number demanded would be three. The labor demand curve that summarizes the store's profit-maximizing employment of detectives is shown in Figure 3.2.

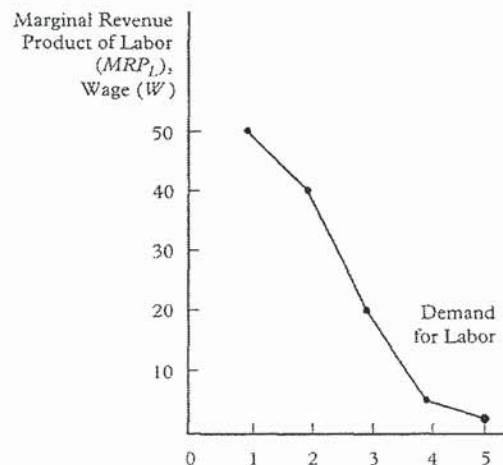
Figure 3.2 illustrates a fundamental point: the labor demand curve in the short run slopes downward because it is the  $MRP_L$  curve—and the  $MRP_L$  curve slopes downward because of labor's diminishing marginal product. The demand curve and the  $MRP_L$  curve coincide; this could be demonstrated by graphing the  $MRP_L$  schedule in Table 3.2, which would yield exactly the same curve as in Figure 3.2. When one detective is hired,  $MRP_L$  is \$50; when two are hired,  $MRP_L$  is \$40; and so forth. Since  $MRP_L$  always equals  $W$  for a profit maximizer who takes wages as given, the  $MRP_L$  curve and labor demand curve (expressed as a function of the money wage) must be the same.

An implication of our example is that there is some level of shoplifting the store finds more profitable to tolerate than to eliminate. This level will be higher at high wages for store detectives than at lower wages. To say the theft rate is "too low" thus implies that the marginal costs of crime reduction exceed the marginal savings generated, and the firm is therefore failing to maximize profits.

Finally, we must emphasize that the marginal product of an individual is *not* a function solely of his or her personal characteristics. As stressed above, the

**FIGURE 3.2**

Demand for Labor in the Short Run (Money Wage)



marginal product of a worker depends upon the number of similar employees the firm has already hired. An individual's marginal product also depends upon the size of the firm's capital stock; increases in the firm's capital stock shift the entire marginal product of labor schedule up. It is therefore incorrect to speak of an individual's productivity as an immutable factor that is associated only with his or her characteristics, independent of the characteristics of the other inputs he or she has to work with.

**MARKET DEMAND CURVES** The demand curve (or schedule) for an individual firm indicates how much labor that firm will want to employ at each wage level. A market demand curve (or schedule) is just the summation of the labor demanded by all firms in a particular labor market at each level of the real wage.<sup>5</sup> If there are three firms in a certain labor market, and if at a given real wage firm A wants 12 workers, firm B wants 6, and firm C wants 20, then the market demand at that real wage is 38 employees. More important, because market demand curves are so closely derived from firm demand curves, they too will slope downward as a function of the real wage. When the real wage falls, the number of workers that existing firms want to employ increases. In addition, the lower real wage may make it profitable for new firms to enter the market. Conversely, when the real wage increases, the number of workers that existing firms want to employ decreases, and some firms may be forced to cease operations completely.

**OBJECTIONS TO THE MARGINAL PRODUCTIVITY THEORY OF DEMAND** Two kinds of objections are sometimes raised to the theory of labor demand introduced in this section. The first is that almost no employer can ever be heard uttering the words "marginal revenue product of labor," and that the theory assumes a degree of sophistication that most employers do not have. Employers, it is also argued, are unable in many situations to accurately measure the output of individual workers.

These first objections can be answered as follows: Whether employers can verbalize the profit-maximizing conditions, or whether they can explicitly measure the marginal revenue product of labor, they must at least *intuit them to survive* in a competitive environment. Competition will "weed out" employers who are not good at generating profits, just as competition will weed out pool players who do not understand the intricacies of how speed, angles, and spin affect the motion of bodies through space. Yet one could canvass the pool halls of America and probably not find one player who could verbalize Newton's laws of motion! The point is that employers can *know* concepts without being able to verbalize them. Those that are not good at maximizing profits will not last very long in competitive markets.

<sup>5</sup>If firms' demand curves are drawn as a function of the money wage, they represent the downward-sloping portion of the firms' marginal revenue product of labor curves. In a competitive industry, the price of the product is given to the firm by the market, and thus at the firm level the marginal revenue product of labor has imbedded in it a given product price. When aggregating labor demand to the market level, product price can no longer be taken as given, and the aggregation is no longer a simple summation.

The second objection is that in many cases, it seems that adding labor while holding capital constant would not add to output at all. For example, one secretary and one computer can produce output, but it might seem that adding a second secretary while holding the number of word processors constant could produce nothing extra, since that secretary would have no machine on which to work.

The answer to this second objection is that the two secretaries could take turns using the computer, so that neither became fatigued to the extent that mistakes increased and typing speeds slowed down. The second secretary could also answer the telephone and in other ways expedite work. Thus, even with technologies that seem to require one machine per person, labor will generally have a marginal product greater than zero if capital is held constant.

## The Demand for Labor in Competitive Markets When Other Inputs Can Be Varied

An implication of our theory of labor demand is that, because labor can be varied in the short run—that is, at any time—the profit-maximizing firm will always operate so that labor's marginal revenue product equals the wage rate (which is labor's marginal expense in a competitive labor market). What we must now consider is how the firm's ability to adjust *other* inputs affects the demand for labor. We first analyze the implications of being able to adjust capital in the long run, and we then turn our attention to the case of more than two inputs.

### Labor Demand in the Long Run

To maximize profits in the long run, the firm must adjust both labor and capital so that the marginal revenue product of each equals its marginal expense. Using the definitions discussed earlier in this chapter, profit maximization requires that the following two equalities be satisfied:

$$MP_L \cdot P = W \quad (\text{a restatement of equation 3.5}) \quad (3.7a)$$

$$MP_K \cdot P = C \quad (\text{the profit-maximizing condition for capital}) \quad (3.7b)$$

Both (3.7a) and (3.7b) can be rearranged to isolate  $P$ , so these two profit-maximizing conditions also can be expressed as

$$P = W/MP_L \quad (\text{a rearrangement of equation 3.7a}) \quad (3.8a)$$

$$P = C/MP_K \quad (\text{a rearrangement of equation 3.7b}) \quad (3.8b)$$

Further, because the right-hand sides of both (3.8a) and (3.8b) equal the same quantity,  $P$ , profit maximization therefore requires that

$$W/MP_L = C/MP_K \quad (3.8c)$$

The economic meaning of equation (3.8c) is key to understanding how the ability to adjust capital affects the firm's demand for labor. Consider the left-hand side of (3.8c): the numerator is the cost of a unit of labor, while the denominator is the extra output produced by an added unit of labor. Therefore, the ratio  $W/MP_L$  turns out to be the added cost of producing an added unit of output when using labor to generate the increase in output.<sup>6</sup> Analogously, the right-hand side is the marginal cost of producing an extra unit of output using capital. What equation (3.8c) suggests is that, to maximize profits, the firm must adjust its labor and capital inputs so that the marginal cost of producing an added unit of output using labor is equal to the marginal cost of producing an added unit of output using capital. Why is this condition a requirement for maximizing profits?

To maximize profits a firm must be producing its chosen level of output in the least-cost manner. Logic suggests that as long as the firm can expand output more cheaply using one input than the other, it cannot be producing in the least-cost way. For example, if the marginal cost of expanding output by one unit using labor were \$10, and the marginal cost using capital were \$12, the firm could keep output constant and lower its costs of production! How? It could reduce its capital by enough to cut output by one unit (saving \$12), and then add enough labor to restore the one-unit cut (costing \$10). Output would be the same, but costs would have fallen by \$2. Thus, for the firm to be maximizing profits, it must be operating at the point such that further marginal changes in both labor and capital would neither lower costs nor otherwise add to profits.

With equations (3.8a) to (3.8c) in mind, what would happen to the demand for labor in the long run if the wage rate ( $W$ ) facing a profit-maximizing firm were to rise? First, as we discussed in the section on the short-run demand for labor, the rise in  $W$  disturbs the equality in (3.8a), and the firm will want to cut back on its use of labor even before it can adjust capital. Because the marginal product of labor is assumed to rise as employment is reduced, any cuts in labor will raise  $MP_L$ .

Second, because each unit of capital now has less labor working with it, the marginal product of capital ( $MP_K$ ) falls, disturbing the equality in (3.8b). By itself, this latter inequality will cause the firm to want to reduce its stock of capital.

Third, the rise in  $W$  will initially end the equality in (3.8c), meaning that the marginal cost of production using labor now exceeds the marginal cost using capital. If the above cuts in labor are made in the short run, the associated increase

<sup>6</sup>Because  $MP_L = \Delta Q / \Delta L$ , the expression  $W/MP_L$  can be rewritten as  $W \cdot \Delta L / \Delta Q$ . Since  $W \cdot \Delta L$  represents the added cost from employing one more unit of labor, the expression  $W \cdot \Delta L / \Delta Q$  equals the cost of an added unit of output when that unit is produced by adding labor.

**EXAMPLE 3.2****Coal Mining Wages and Capital Substitution**

That wage increases have both a *scale effect* and a *substitution effect*, both of which tend to reduce employment, is widely known—even by many of those pushing for higher wages. John L. Lewis was president of the United Mine Workers from the 1920s through the 1940s, when wages for miners were increased considerably with full knowledge that this would induce the substitution of capital for labor. According to Lewis:

Primarily the United Mine Workers of America insists upon the maintenance of the wage standards guaranteed by the existing contractual relations in the industry, in the interests of its own membership.... But in insisting on the maintenance of an American wage standard in the coal fields the United Mine Workers is also doing its

part, probably more than its part, to force a reorganization of the basic industry of the country upon scientific and efficient lines. The maintenance of these rates will accelerate the operation of natural economic laws, which will in time eliminate uneconomic mines, obsolete equipment, and incompetent management.

The policy of the United Mine Workers of America will inevitably bring about the utmost employment of machinery of which coal mining is physically capable.... Fair wages and American standards of living are inextricably bound up with the progressive substitution of mechanical for human power. It is no accident that fair wages and machinery will walk hand-in-hand.

Source: John L. Lewis, *The Miners' Fight for American Standards* (Indianapolis: Bell, 1925), 40, 41, 108.

in  $MP_L$  and decrease in  $MP_K$  will work toward restoring equality in (3.8c); however, if it remains more costly to produce an extra unit of output using labor than using capital, the firm will want to substitute capital for labor in the long run. Substituting capital for labor means that the firm will produce its profit-maximizing level of output (which is clearly reduced by the rise in  $W$ ) in a more capital-intensive way. The act of substituting capital for labor also will serve to increase  $MP_L$  and reduce  $MP_K$ , thereby reinforcing the return to equality in (3.8c).

In the end, the increase in  $W$  will cause the firm to reduce its desired employment level for two reasons. The firm's profit-maximizing level of output will fall, and the associated reduction in required inputs (both capital and labor) is an example of the *scale effect*. The rise in  $W$  also causes the firm to substitute capital for labor, so that it can again produce in the least-cost manner; changing the mix of capital and labor in the production process is an example of the *substitution effect*. The scale and substitution effects of a wage increase will have an ambiguous effect on the firm's desired stock of capital, but both effects serve to reduce the demand for labor. Thus, as illustrated in Example 3.2, the long-run ability to adjust capital lends further theoretical support to the proposition that the labor demand curve is a downward-sloping function of the wage rate.

**More than Two Inputs**

Thus far we have assumed that there are only two inputs in the production process: capital and labor. In fact, labor can be subdivided into many categories; for exam-

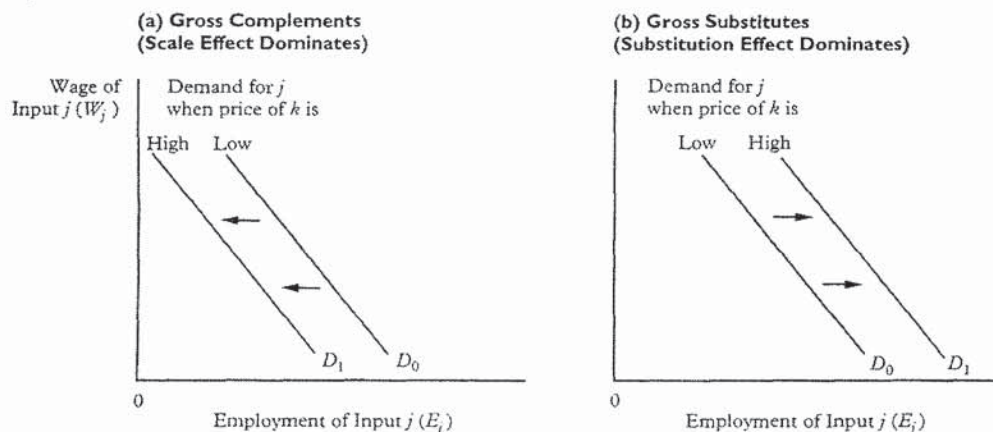
ple, labor can be categorized by age, educational level, and occupation. Other inputs that are used in the production process include materials and energy. If a firm is seeking to minimize costs, in the long run it should employ all inputs up until the point that the marginal cost of producing an added unit of output is the same regardless of which input is increased. This generalization of equation (3.8c) leads to the somewhat obvious result that the demand for any category of labor will be a function of its own wage rate and (through the scale and substitution effects) the wage or prices of all other categories of labor, capital, and supplies.

**IF INPUTS ARE SUBSTITUTES IN PRODUCTION** The demand curve for each category of labor will be a downward-sloping function of the wage rate paid to workers in that category for the reasons discussed earlier, but how is it affected by wage or price changes for *other* inputs? If two inputs are substitutes in production (that is, if the greater use of one in producing output can compensate for reduced use of the other), then increases in the price of the *other* input may shift the entire demand curve for a *given* category of labor either to the right or to the left, depending on the relative strength of the substitution and scale effects. If an increase in the price of one input shifts the demand for *another* input to the left, as in panel (a) of Figure 3.3, the scale effect has dominated the substitution effect and the two inputs are said to be *gross complements*; if the increase shifts the demand for the other input to the right, as in panel (b) of Figure 3.3, the substitution effect has dominated and the two inputs are *gross substitutes*.

**IF INPUTS ARE COMPLEMENTS IN PRODUCTION** If, instead, the two inputs must be used together—in which case they are called *perfect complements* or *complements in production*—then reduced use of one implies reduced use of the other.

FIGURE 3.3

Effect of Increase in the Price of One Input ( $k$ ) on Demand for Another Input ( $j$ ), where Inputs Are Substitutes in Production



In this case, there is no substitution effect, only a scale effect, and the two inputs must be gross complements.

**EXAMPLES** Consider an example of a snow-removal firm in which skilled and unskilled workers are substitutes in production—snow can be removed using either unskilled workers (with shovels) or skilled workers driving snowplows. Let us focus on demand for the skilled workers. Other things equal, an increase in the wage of skilled workers would cause the firm to employ fewer of them; their demand curve would be a downward-sloping function of their wage. If only the wage of *unskilled* workers increased, however, the employer would want fewer unskilled workers than before, and more of the now relatively less expensive skilled workers, to remove any *given amount of snow*. To the extent that this substitution effect dominated over the scale effect, the demand for skilled workers would shift to the right. In this case, skilled and unskilled workers would be gross substitutes. In contrast, if the reduction in the scale of output caused employment of skilled workers to be reduced, even though skilled workers were being substituted for unskilled workers in the production process, skilled and unskilled workers would be considered gross complements.

In the above firm, snowplows and skilled workers are complements in production. If the price of snowplows went up, the employer would want to cut back on their use, which would result in a reduced demand at each wage for the skilled workers who drove the snowplows. As noted above, inputs that are complements in production are always gross complements.

## Labor Demand When the Product Market Is Not Competitive

Our analysis of the demand for labor, in both the short and the long run, has so far taken place under the assumption that the firm operates in competitive product and labor markets. This is equivalent to assuming that the firm is both a price taker and a wage taker; that is, that it takes both  $P$  and  $W$  as given and makes decisions only about the levels of output and inputs. We will now explore the effects of noncompetitive (monopolistic) product markets on the demand for labor (the effects of noncompetitive labor markets will be analyzed in chapter 5).

### Maximizing Monopoly Profits

As explained earlier in footnote 2 and the surrounding text, product market monopolies are subject to the market demand curve for their output, and they therefore do not take output price as given. They can expand their sales only by reducing product price, which means that their marginal revenue ( $MR$ ) from an extra unit of output is less than product price ( $P$ ). Using the general definition of marginal revenue product in equation (3.3a), and applying the usual profit-maximizing criteria outlined in equation (3.4) to a monopoly that searches for workers in a

competitive *labor* market (so that  $ME_L = W$ ), the monopolist would hire workers until its marginal revenue product of labor ( $MRP_L$ ) equals the wage rate:

$$MRP_L = MR \cdot MP_L = W \quad (3.9)$$

Now we can express the demand for labor in the short run in terms of the real wage by dividing equation (3.9) by the firm's product price,  $P$ , to obtain

$$\frac{MR}{P} \cdot MP_L = \frac{W}{P} \quad (3.10)$$

Since marginal revenue is always less than a monopoly's product price, the ratio  $MR/P$  in equation (3.10) is less than one. Therefore, the labor demand curve for a firm that has monopoly power in the output market will lie below and to the left of the labor demand curve for an *otherwise identical* firm that takes product price as given. Put another way, just as the level of profit-maximizing output is lower under monopoly than it is under competition, other things equal, so is the level of employment.

The wage rates that monopolies pay, however, are not necessarily different from competitive levels even though employment levels are. An employer with a product market monopoly may still be a very small part of the market for a particular kind of employee, and thus be a *price taker* in the labor market. For example, a local utility company might have a product market monopoly, but it would have to compete with all other firms to hire secretaries and thus would have to pay the going wage.

### Do Monopolies Pay Higher Wages?

There are circumstances, however, in which economists suspect that product market monopolies might pay wages that are *higher* than competitive firms would pay.<sup>7</sup> The monopolies that are legally permitted to exist in the United States are regulated by governmental bodies in an effort to prevent them from exploiting their favored status and earning monopoly profits. This regulation of profits, it can be argued, gives monopolies incentives to pay higher wages than they would otherwise pay, for two reasons.

First, regulatory bodies allow monopolies to pass the costs of doing business on to consumers. Thus, while unable to maximize profits, the managers of a monopoly can enhance their *utility* by paying high wages and passing the costs along to consumers in the form of higher prices. The ability to pay high wages makes a

<sup>7</sup>For a full statement of this argument, see Armen Alchian and Reuben Kessel, "Competition, Monopoly, and the Pursuit of Money," in *Aspects of Labor Economics*, ed. H. G. Lewis (Princeton, N.J.: Princeton University Press, 1962).

manager's life more pleasant by making it possible to hire people who might be more attractive or personable or have other characteristics managers find desirable.

Second, monopolies that are as yet unregulated may not want to attract attention to themselves by earning the very high profits usually associated with monopoly. Therefore they, too, may be induced to pay high wages in a partial effort to "hide" their profits. The excess profits of monopolies, in other words, may be partly taken in the form of highly preferred workers—paid a relatively high wage rate—rather than in the usual monetary form.

The evidence on monopoly wages, however, is not very clear as yet. Some studies suggest that firms in industries with relatively few sellers *do* pay higher wages than competitive firms for workers with the same education and experience. Other studies of regulated monopolies, however, have obtained mixed results on whether wages tend to be higher for comparable workers in these industries.<sup>8</sup>

## Policy Application: The Labor Market Effects of Employer Payroll Taxes and Wage Subsidies

We now turn to an application of labor demand theory to the phenomena of employer payroll taxes and wage subsidies. Governments widely finance certain social programs through taxes that require *employers* to remit payments based on their total payroll costs. As we will see, new or increased payroll taxes levied on the employer raise the cost of hiring labor, and they might therefore be expected to reduce the demand for labor. Conversely, it can be argued that if the government were to subsidize the wages paid by employers, the demand for labor would increase; indeed, wage subsidies for particular disadvantaged groups in society are sometimes proposed as a way to increase their employment. In this section, we will analyze the effects of payroll taxes and subsidies.

### Who Bears the Burden of a Payroll Tax?

Payroll taxes require employers to pay the government a certain percentage of their employees' earnings, often up to some maximum amount. Unemployment insurance, as well as Social Security retirement, disability, and Medicare programs, are prominent examples. Does taxing employers to generate revenues for these programs relieve *employees* of a financial burden that would otherwise fall on them?

Suppose that only the employer is required to make payments and that the tax is a fixed amount ( $X$ ) per labor hour, rather than a percentage of payroll. Now

<sup>8</sup>Ronald Ehrenberg, *The Regulatory Process and Labor Earnings* (New York: Academic Press, 1979); Barry T. Hirsch, "Trucking Regulation, Unionization, and Labor Earnings," *Journal of Human Resources* 23 (Summer 1988): 296–319; S. Nickell, J. Vainiomaki, and S. Wadhvani, "Wages and Product Market Power," *Economica* 61 (November 1994): 457–473; and Marianne Bertrand and Sendhil Mullainathan, "Is There Discretion in Wage Setting? A Test Using Takeover Legislation," *RAND Journal of Economics* 30 (Autumn 1999): 535–554.

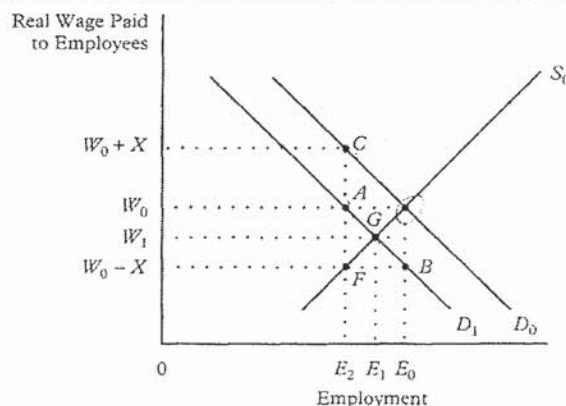
consider the market demand curve  $D_0$  in Figure 3.4, which is drawn in such a way that desired employment is plotted against the wage *employees receive*. Prior to the imposition of the tax, the wage employees receive is the same as the wage employers pay. Thus, if  $D_0$  were the demand curve before the tax was imposed, it would have the conventional interpretation of indicating how much labor firms would be willing to hire at any given wage. However, *after* imposition of the tax, employer wage costs would be  $X$  above what employees received.

**SHIFTING THE DEMAND CURVE** If employees received  $W_0$ , employers would now face costs of  $W_0 + X$ . They would no longer demand  $E_0$  workers; rather, because their costs were  $W_0 + X$ , they would demand  $E_2$  workers. Point  $A$  (where  $W_0$  and  $E_2$  intersect) would lie on a new market demand curve, formed when demand shifted down because of the tax (remember, the wage on the vertical axis of Figure 3.4 is the wage *employees receive*, not the wage employers pay). Only if employee wages fell to  $W_0 - X$  would firms want to continue hiring  $E_0$  workers, for then employer costs would be the same as before the tax. Thus, point  $B$  would also be on the new, shifted demand curve. Note that, with a tax of  $X$ , the new demand curve ( $D_1$ ) is parallel to the old one and the vertical distance between the two is  $X$ .

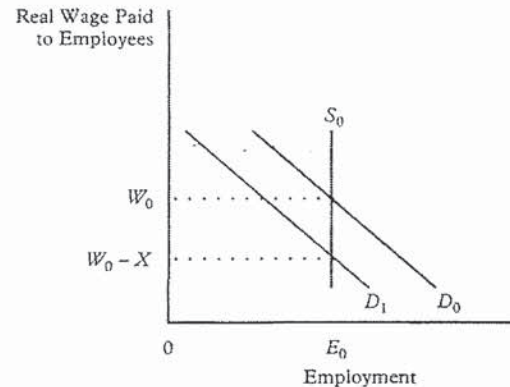
Now, the tax-related shift in the market demand curve to  $D_1$  implies that there would be an excess supply of labor at the previous equilibrium wage of  $W_0$ . This surplus of labor would create downward pressure on the *employee* wage, and this downward pressure would continue to be exerted until the employee wage fell to  $W_1$ , the point at which the quantity of labor supplied just equaled the quantity demanded. At this point, employment would also have fallen to  $E_1$ . Thus, *employees bear a burden in the form of lower wage rates and lower employment levels. The lesson is clear: employees are not exempted from bearing costs when the government chooses to generate revenues through a payroll tax on employers.*

Figure 3.4 does suggest, however, that employers may bear at least some of the tax, because the wages received by employees do not fall by the full amount of the tax ( $W_0 - W_1$  is smaller than  $X$ , which is the vertical distance between the

**FIGURE 3.4**  
The Market Demand Curve and Effects of an Employer-Financed Payroll Tax



**FIGURE 3.5**  
Payroll Tax with a Vertical Supply Curve



two demand curves). This occurs because, with an upward-sloping labor market supply curve, employees withdraw labor as their wages fall, and it becomes more difficult for firms to find workers. If wages fell to  $W_0 - X$ , the withdrawal of workers would create a labor shortage that would drive wages to some point ( $W_1$  in our example) between  $W_0$  and  $W_0 - X$ . Only if the labor market supply curve were *vertical*—meaning that lower wages have no effect on labor supply—would the *entire amount of the tax* be shifted to workers in the form of a decrease in their wages by the amount of  $X$  (see Figure 3.5).

**EFFECTS OF LABOR SUPPLY CURVES** The extent to which the labor market supply curve is sensitive to wages affects the proportion of the employer payroll tax that gets shifted to employees' wages. The less responsive labor supply is to changes in wages, the fewer the employees who withdraw from the market and the higher the proportion of the tax that gets shifted to workers in the form of a wage decrease (compare the outcomes in Figures 3.4 and 3.5). It must also be pointed out, however, that to the degree employee wages do *not* fall, employment levels *will*; when employee wages do not fall much in the face of an employer payroll-tax increase, employer labor costs are increased—and this increase reduces the quantity of labor employers demand.

A number of empirical studies have sought to ascertain what fraction of employers' payroll-tax costs are actually passed on to employees in the form of lower wages (or lower wage increases). Although the evidence is somewhat ambiguous, a comprehensive review of these studies led to at least a tentative conclusion that most of a payroll tax is eventually shifted to wages, with little long-run effect on employment.<sup>9</sup>

<sup>9</sup>Daniel S. Hamermesh, *Labor Demand* (Princeton, N.J.: Princeton University Press, 1993), 169–173. Also see Jonathan Gruber, "The Incidence of Payroll Taxation: Evidence from Chile," *Journal of Labor Economics* 15, no. 3, pt. 2 (July 1997): S72–S101; Patricia M. Anderson and Bruce D. Meyer, "The Effects of the Unemployment Insurance Payroll Tax on Wages, Employment, Claims and Denials," *Journal of Public Economics* 78 (October 2000): 81–106; and Kevin Lang, "The Effect of the Payroll Tax on Earnings: A Test of Competing Models of Wage Determination," National Bureau of Economic Research Working Paper no. 9537, February 2003.

### Employment Subsidies as a Device to Help the Poor

The opposite of a payroll tax on employers is a government subsidy of employers' payrolls. In Figure 3.4, for example, if instead of *taxing* each hour of labor by  $X$  the government *paid* the employer  $X$ , the market labor demand curve would shift *upward* by a vertical distance of  $X$ . This upward movement of the demand curve would create pressures to increase employment and the wages received by employees; as with a payroll tax, whether the eventual effects would be felt more on employment or on wage rates depends on the shape of the labor market supply curve.

(Students should test their understanding in this area by drawing labor demand curves that reflect a new payroll subsidy of  $X$  per hour, and then analyzing the effects on employment and employee wages with market supply curves that are, alternatively, upward-sloping and vertical. *Hint:* The outcomes should be those that would be obtained if demand curve  $D_1$  in Figures 3.4 and 3.5 were shifted by the subsidy to curve  $D_0$ .)

Payroll subsidies to employers can take many forms. They can be in the form of cash payments, as implied by the above hypothetical example, or they can be in the form of tax credits. These credits might directly reduce a firm's payroll-tax rate, or they might reduce some other tax by an amount proportional to the number of labor hours hired; in either case, the credit has the effect of reducing the cost of hiring labor.

Further, wage subsidies can apply to a firm's employment *level*, to any *new* employees hired after a certain date (even if they just replace workers who have left), or only to new hires that serve to *increase* the firm's level of employment. Finally, subsidies can be either *general* or *selective*. A general subsidy is not conditional on the characteristics of the people hired, whereas a selective, or *targeted*, plan makes the subsidy conditional on hiring people from certain target groups (such as the disadvantaged).

The beneficial effects on wages and employment that are expected to derive from payroll-tax subsidies have led to their proposed use as a policy to help alleviate poverty. As one economist concerned about unemployment and earnings levels among low-wage workers wrote:

What to do? The solution for which I have pleaded the past five years: a low-wage employment subsidy. It would best take the form of a tax credit that employers could use to offset the payroll taxes they owe from their employment of low-wage workers. Lower unemployment and better pay would result at the low end of the labor market—the less of the one, the more of the other.<sup>10</sup>

Experience in the United States with targeted wage subsidies, such as the one proposed above, has been modest. The Targeted Jobs Tax Credit (TJTC) program, which began in 1979 and was changed slightly over the years until it was finally

<sup>10</sup>Edmund S. Phelps, "Commentary: Past and Prospective Causes of High Unemployment," in *Reducing Unemployment: Current Issues and Policy Options* (Kansas City, Mo.: Federal Reserve Bank of Kansas City, 1994), 89. For his book on the topic, see Edmund S. Phelps, *Rewarding Work* (Cambridge, Mass.: Harvard University Press, 1997).

## EMPIRICAL STUDY

## Do Women Pay for Employer-Funded Maternity Benefits? Using Cross-Section Data over Time to Analyze “Differences in Differences”

During the last half of 1976, Illinois, New Jersey, and New York passed laws requiring that employer-provided health insurance plans treat pregnancy the same as illness (that is, coverage of doctor's bills and hospital costs had to be the same for pregnancy as for illnesses or injuries). These mandates increased the cost of health insurance for women of childbearing age by an amount that was equal to about 4 percent of their earnings. Were these increases in employer costs borne by employers, or did they reduce the wages of women by an equivalent amount?

A problem confronting researchers on this topic is that the adopting states are all states with high incomes and likely to have state legislation encouraging the expansion of employment opportunities for women. Thus, comparing wage *levels* across states would require that we statistically control for all the factors, besides the maternity-benefit mandate, that affect wages. Because we can never be sure that we have adequate controls for the economic, social, and legal factors that affect wage levels by state, we need to find another way to perform the analysis.

Fortunately, answering the research question is facilitated by several factors: (a) some states adopted these laws and some did not; (b) even in states that

adopted these laws, the insurance cost increases applied only to women (and their husbands) of childbearing age, and not to single men or older workers; and (c) the adopting states passed these laws during the same time period, so that variables affecting women's wages that change over time (such as recessions or the rising presence of women in the labor force) do not cloud the analysis.

Factors (a) and (c) above allow the conduct of what economists call a “differences-in-differences” analysis. Specifically, these factors allow us to compare wage *changes*, from the pre-adoption years to the post-adoption ones, among women of childbearing age in adopting states (the “experimental group”) to wage changes over the same period for women of the same age in states that did not adopt (a “comparison group”). By comparing within-state *changes* in wages, we avoid the need to find measures that would control for the economic, social, and public-policy forces that make the initial wage *level* in one state differ from that in another; whatever the factors are that raise wage levels in New Jersey, for example, they were there both before and after the adoption of mandated maternity benefits.

One might argue, of course, that the adopting and nonadopting states were

subject to *other* forces (unrelated to maternity benefits) that led to different degrees of wage change over this period. For example, the economy of New Jersey might have been booming during the period when maternity benefits were adopted, while economies elsewhere might not have been. However, if an adopting state is experiencing unique wage pressures in addition to those imposed by maternity benefits, the effects of these other pressures should show up in the wage changes experienced by single men or older women—groups in the adopting states that were not affected by the mandate. Thus, we can exploit factor (b) above by also comparing the wage changes for women of childbearing age in adopting states to those for single men or older women in the same states.

The three factors above enabled one researcher to measure how the wages of married women, age 20–40, changed from

1974–1975 to 1977–1978 in the three adopting states. These changes were then compared to changes in wages for married women of the same age in non-adopting (but economically similar) states. To account for forces *other than changing maternity benefits* that could affect wage changes across states during this period, the researcher also measured changes in wages for unmarried men and workers over 40 years of age. This study concluded that in the states adopting mandated maternity benefits, the post-adoption wages of women in the 20–40 age group were about 4 percent lower than they would have been without adoption. This finding suggests that the entire cost of maternity benefits was quickly shifted to women of childbearing age.

Source: Jonathan Gruber, "The Incidence of Mandated Maternity Benefits," *American Economic Review* 84 (June 1994): 622–641.

discontinued in 1995, targeted disadvantaged youth, the handicapped, and welfare recipients, providing their employers with a tax credit that lasted for one year. In practice, the average duration of jobs under this program was six months, and the subsidy reduced employer wage costs by about 15 percent for jobs of this duration.

One problem that limited the effectiveness of the TJTC program was that the eligibility requirements for many of its participants were stigmatizing; that is, being eligible (on welfare, for example) was often seen by employers as a negative indicator of productivity. Nevertheless, one evaluation found that the employment of disadvantaged youth was enhanced by the TJTC. Specifically, it found that when 23- to 24-year-olds were removed from eligibility for the TJTC by changes in 1989, employment of disadvantaged youths of that age fell over 7 percent.<sup>11</sup>

<sup>11</sup>Lawrence F. Katz, "Wage Subsidies for the Disadvantaged," in *Generating Jobs: How to Increase Demand for Less-Skilled Workers*, ed. Richard B. Freeman and Peter Gottschalk (New York: Russell Sage Foundation, 1998), 21–53.

## Review Questions

1. In a statement during the 1992 presidential campaign, one organization attempting to influence the political parties argued that the wages paid by U.S. firms in their Mexican plants were so low that they “have no relationship with worker productivity.” Comment on this statement using the principles of profit maximization.
2. Assume that wages for keyboarders (data entry clerks) are lower in India than in the United States. Does this mean that keyboarding jobs in the United States will be lost to India? Explain.
3. The Occupational Safety and Health Administration promulgates safety and health standards. These standards typically apply to machinery (capital), which is required to be equipped with guards, shields, and the like. An alternative to these standards is to require the employer to furnish personal protective devices to employees (labor)—such as earplugs, hard hats, and safety shoes. *Disregarding* the issue of which alternative approach offers greater protection from injury, what aspects of each alternative must be taken into account when analyzing the possible *employment* effects of the two approaches to safety?
4. Suppose that prisons historically have required inmates to perform, *without pay*, various cleaning and food preparation jobs within the prison. Now suppose that prisoners are offered paid work in factory jobs within the prison walls, and that the cleaning and food preparation tasks are now performed by nonprisoners hired to do them. Would you expect to see any differences in the *technologies* used to perform these tasks? Explain.
5. Years ago, Great Britain adopted a program that placed a tax—to be collected from employers—on wages in *service* industries. Wages in manufacturing industries were not taxed. Discuss the wage and employment effects of this tax policy.
6. Suppose the government were to subsidize the wages of all women in the population by paying their *employers* 50 cents for every hour they work. What would be the effect on the wage rate women received? What would be the effect on the net wage employers paid? (The net wage would be the wage women received less 50 cents.)
7. In the last two decades, the United States has been subject to huge increases in the illegal immigration of workers from Mexico, most of them unskilled, and the government has considered ways to reduce the flow. One policy is to impose larger financial penalties on employers who are discovered to have hired illegal immigrants. What effect would this policy have on the employment of unskilled illegal immigrants? What effect would it have on the demand for skilled “native” labor?
8. If anti-sweatshop movements are successful in raising pay and improving working conditions for apparel workers in foreign countries, how will these changes abroad affect labor market outcomes for workers in the apparel and retailing industries in the United States? Explain.

## Problems

1. An experiment conducted in Tennessee found that the scores of second and third graders on standardized tests for reading, math, listening, and word study skills were the same in small classrooms (13 to 17 students) as in regular classrooms (22 to 25 students). Suppose that there is a school that had 90 third graders taught by four teachers and that added two additional teachers to reduce the class size. If the Tennessee study can be generalized, what is the marginal product of labor of these two additional teachers?
2. The marginal revenue product of labor in the local sawmill is  $MRP_L = 20 - 0.5L$ , where  $L$  = the number of workers. If the wage of sawmill workers is \$10 per hour, then how many workers will the mill hire?
3. Suppose that the supply curve for lifeguards is  $L_S = 20$  and the demand curve for lifeguards is  $L_D = 100 - 20W$ , where  $L$  = the number of lifeguards and  $W$  = the hourly wage. Graph both the demand and supply curves. Now suppose that the government imposes a tax of \$1 per hour per worker on companies hiring lifeguards. Draw the new (after-tax) demand curve in terms of the employee wage. How will this tax affect the wage of lifeguards and the number employed as lifeguards?
4. The output of workers at a factory depends on the number of supervisors hired (see table below). The factory sells its output for \$0.50 each, it hires 50 production workers at a wage of \$100 per day, and it needs to decide how many supervisors to hire. The daily wage of supervisors is \$500, but output rises as more supervisors are hired, as shown in the table. How many supervisors should it hire?
5. (Appendix) The Hormsburry Corporation produces yo-yos at its factory. Both its labor and capital markets are competitive. Wages are \$12 per hour and yo-yo-making equipment (a computer-controlled plastic extruding machine) rents for \$4 per hour. The production function is  $q = 40K^{0.25}L^{0.75}$ , where  $q$  = boxes of yo-yos per week,  $K$  = hours of yo-yo equipment used, and  $L$  = hours of labor. Therefore,  $MP_L = 30K^{0.25}L^{-0.25}$  and  $MP_K = 10K^{-0.75}L^{0.75}$ . Determine the cost-minimizing capital-labor ratio at this firm.

Supervisors	Output (units per day)
0	11,000
1	14,800
2	18,000
3	19,500
4	20,200
5	20,600

## Selected Readings

Blank, Rebecca M., ed. *Social Protection versus Economic Flexibility: Is There a Trade-Off?* Chicago: University of Chicago Press, 1994.

Hamermesh, Daniel. *Labor Demand*. Princeton, N.J.: Princeton University Press, 1993.

Katz, Lawrence F. "Wage Subsidies for the Disadvantaged." In *Generating Jobs: How to Increase Demand for Less-Skilled Workers*, ed. Richard B. Freeman and Peter Gottschalk, 21–53. New York: Russell Sage Foundation, 1998.



## APPENDIX 3A

# Graphical Derivation of a Firm's Labor Demand Curve

Chapter 3 described verbally the derivation of a firm's labor demand curve. This appendix will present the *same* derivation graphically. This graphical representation permits a more rigorous derivation, but our conclusion that demand curves slope downward in both the short and the long run will remain unchanged.

## The Production Function

Output can generally be viewed as being produced by combining capital and labor. Figure 3A.1 illustrates this production function graphically and depicts several aspects of the production process.

Consider the convex curve labeled  $Q = 100$ . Along this line, every combination of labor ( $L$ ) and capital ( $K$ ) produces 100 units of output ( $Q$ ). That is, the combination of labor and capital at point  $A (L_a, K_a)$  generates the same 100 units of output as the combinations at points  $B$  and  $C$ . Because each point along the  $Q = 100$  curve generates the same output, that curve is called an *isoquant* (*iso* = "equal"; *quant* = "quantity").

Two other isoquants are shown in Figure 3A.1 ( $Q = 150$ ,  $Q = 200$ ). These isoquants represent higher levels of output than the  $Q = 100$  curve. The fact that these isoquants indicate higher output levels can be seen by holding labor constant at  $L_b$  (say) and then observing the different levels of capital. If  $L_b$  is combined with  $K_b$  in capital, 100 units of  $Q$  are produced. If  $L_b$  is combined with  $K'_b$ , 150 units are produced ( $K'_b$  is greater than  $K_b$ ). If  $L_b$  is combined with even more capital ( $K''_b$ , say), 200 units of  $Q$  could be produced.

Note that the isoquants in Figure 3A.1 have *negative* slopes, reflecting an assumption that labor and capital are substitutes. If, for example, we cut capital from  $K_a$  to  $K_b$ , we could keep output constant (at 100) by increasing labor from  $L_a$  to  $L_b$ . Labor, in other words, could be substituted for capital to maintain a given production level.

Finally, note the *convexity* of the isoquants. At point A, the  $Q = 100$  isoquant has a steep slope, suggesting that to keep  $Q$  constant at 100, a given decrease in capital could be accompanied by a *modest* increase in labor. At point C, however, the slope of the isoquant is relatively flat. This flatter slope means that the same given decrease in capital would require a much *larger* increase in labor for output to be held constant. The decrease in capital permitted by a given increase in labor while output is being held constant is called the *marginal rate of technical substitution* (MRTS) between capital and labor. Symbolically, the MRTS can be written as

$$MRTS = \frac{\Delta K}{\Delta L} \Big| \bar{Q} \quad (3A.1)$$

where  $\Delta$  means “change in” and  $\bar{Q}$  means “holding output constant.” The MRTS is *negative*, because if  $L$  is increased,  $K$  must be reduced to keep  $Q$  constant.

Why does the absolute value of the marginal rate of technical substitution diminish as labor increases? When labor is highly used in the production process and capital is not very prevalent (point C in Figure 3A.1), there are many jobs that capital can do. Labor is easy to replace; if capital is increased, it will be used as a substitute for labor in parts of the production process where it will have the highest payoff. As capital becomes progressively more utilized and labor less so, the few remaining workers will be doing jobs that are hardest for a machine to do, at which point it will take a lot of capital to substitute for a worker.<sup>1</sup>

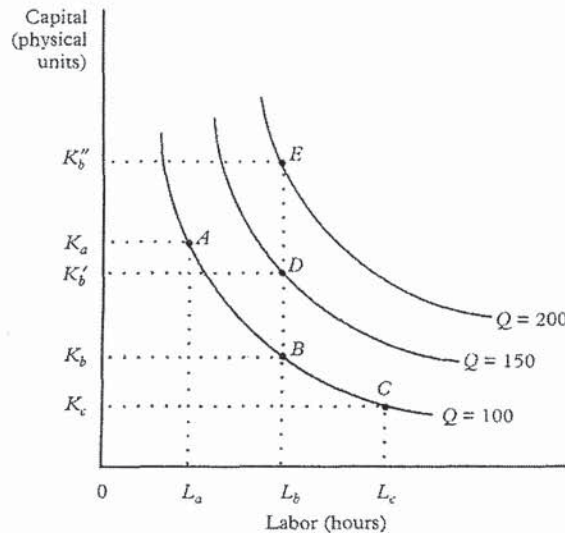
## Demand for Labor in the Short Run

Chapter 3 argued that firms will maximize profits in the short run ( $K$  fixed) by hiring labor until labor's marginal product ( $MP_L$ ) is equal to the real wage ( $W/P$ ). The reason for this decision rule is that the real wage represents the *cost* of an added unit of labor (in terms of output), while the marginal product is the *output* added by the extra unit of labor. As long as the firm, by increasing labor ( $K$  fixed), gains more in output than it loses in costs, it will continue to hire employees. The firm will stop hiring when the marginal cost of added labor exceeds  $MP_L$ .

The requirement that  $MP_L = W/P$  in order for profits to be maximized means that the firm's labor demand curve in the short run (in terms of the *real* wage) is identical to its marginal product of labor schedule (refer to Figure 3.1). Remembering that the marginal product of labor is the extra output produced by one-unit increases in the amount of labor employed, holding capital constant, consider the production

<sup>1</sup>Here is one example. Over time, telephone operators (who used to place long-distance calls) were replaced by a very capital-intensive direct-dialing system. Those operators who remain employed, however, perform tasks that are the most difficult for a machine to perform—handling collect calls, dispensing directory assistance, and acting as troubleshooters when problems arise.

**FIGURE 3A.1**  
A Production Function

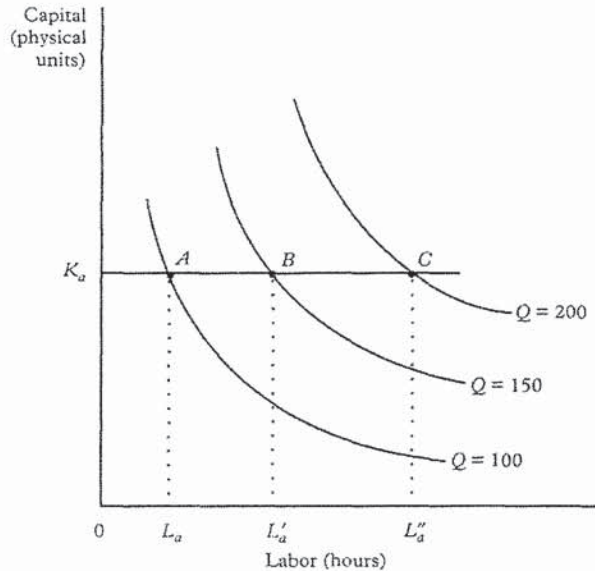


function displayed in Figure 3A.2. Holding capital constant at  $K_a$ , the firm can produce 100 units of  $Q$  if it employs labor equal to  $L_a$ . If labor is increased to  $L'_a$ , the firm can produce 50 more units of  $Q$ ; if labor is increased from  $L'_a$  to  $L''_a$ , the firm can produce an additional 50 units. Notice, however, that the required increase in labor to get the latter 50 units of added output,  $L''_a - L'_a$ , is larger than the extra labor required to produce the first 50-unit increment ( $L'_a - L_a$ ). This difference can only mean that as labor is increased when  $K$  is held constant, each successive labor hour hired generates progressively smaller increments in output. Put differently, Figure 3A.2 graphically illustrates the diminishing marginal productivity of labor.

Why does labor's marginal productivity decline? Chapter 3 explained that labor's marginal productivity declines because, with  $K$  fixed, each added worker has less capital (per capita) with which to work. Is this explanation proven in Figure 3A.2? The answer is, regrettably, no. Figure 3A.2 is drawn *assuming* diminishing marginal productivity. Renumbering the isoquants could produce a different set of marginal productivities. (To see this, change  $Q = 150$  to  $Q = 200$ , and change  $Q = 200$  to  $Q = 500$ . Labor's marginal productivity would then rise.) However, the logic that labor's marginal product must eventually fall as labor is increased, holding buildings, machines, and tools constant, is compelling. Further, as chapter 3 pointed out, even if  $MP_L$  rises initially, the firm will stop hiring labor only in the range where  $MP_L$  is declining; as long as  $MP_L$  is above  $W/P$  and rising, it will pay to continue hiring.

**FIGURE 3A.2**

The Declining Marginal Productivity of Labor



The assumptions that  $MP_L$  declines eventually and that firms hire until  $MP_L = W/P$  are the bases for the assertion that a firm's short-run demand curve for labor slopes downward. The graphical, more rigorous derivation of the demand curve in this appendix confirms and supports the verbal analysis in the chapter. However, it also emphasizes more clearly than a verbal analysis can that the downward-sloping nature of the short-run labor demand curve is based on an *assumption*—however reasonable—that  $MP_L$  declines as employment is increased.

## Demand for Labor in the Long Run

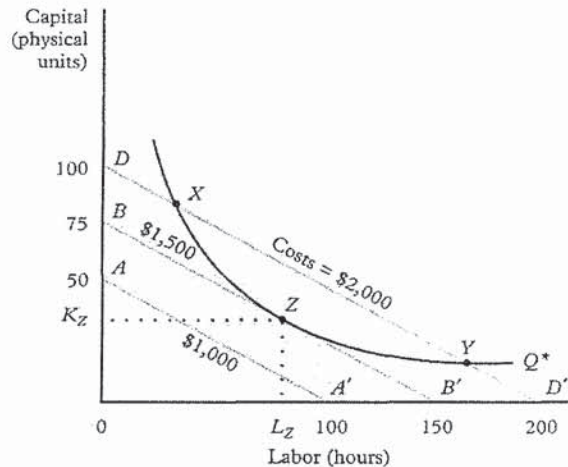
Recall that a firm maximizes its profits by producing at a level of output ( $Q^*$ ) where marginal cost equals marginal revenue. That is, the firm will keep increasing output until the addition to its revenues generated by an extra unit of output just equals the marginal cost of producing that extra unit of output. Because marginal revenue, which is equal to output *price* for a competitive firm, is not shown in our graph of the production function, the profit-maximizing level of output cannot be determined. However, continuing our analysis of the production function can illustrate some important aspects of the demand for labor in the long run.

### Conditions for Cost Minimization

In Figure 3A.3, profit-maximizing output is assumed to be  $Q^*$ . How will the firm combine labor and capital to produce  $Q^*$ ? It can maximize profits only if it produces

**FIGURE 3A.3**

Cost Minimization in the Production of  $Q^*$   
(Wage = \$10 per Hour; Price of a Unit of  
Capital = \$20)



$Q^*$  in the least expensive way; that is, it must minimize the costs of producing  $Q^*$ . To better understand the characteristics of cost minimization, refer to the three *isoexpenditure* lines— $AA'$ ,  $BB'$ ,  $DD'$ —in Figure 3A.3. Along any one of these lines the costs of employing labor and capital are equal.

For example, line  $AA'$  represents total costs of \$1,000. Given an hourly wage ( $W$ ) of \$10 per hour, the firm could hire 100 hours of labor and incur total costs of \$1,000 if it used no capital (point  $A'$ ). In contrast, if the price of a unit of capital ( $C$ ) is \$20, the firm could produce at a total cost of \$1,000 by using 50 units of capital and no labor (point  $A$ ). All the points between  $A$  and  $A'$  represent combinations of  $L$  and  $K$  that, at  $W = \$10$  and  $C = \$20$ , cost \$1,000 as well.

The problem with the isoexpenditure line of  $AA'$  is that it does not intersect the isoquant  $Q^*$ , implying that  $Q^*$  cannot be produced for \$1,000. At prices of  $W = \$10$  and  $C = \$20$ , the firm cannot buy enough resources to produce output level  $Q^*$  and hold total costs to \$1,000. The firm can, however, produce  $Q^*$  for a total cost of \$2,000. Line  $DD'$ , representing expenditures of \$2,000, intersects the  $Q^*$  isoquant at points  $X$  and  $Y$ . The problem with these points, however, is that they are not cost-minimizing;  $Q^*$  can be produced for less than \$2,000.

Since isoquant  $Q^*$  is convex, the cost-minimizing combination of  $L$  and  $K$  in producing  $Q^*$  will come at a point where an isoexpenditure line is *tangent* to the isoquant (that is, just barely touches isoquant  $Q^*$  at only one place). Point  $Z$ , where labor equals  $L_Z$  and capital equals  $K_Z$ , is where  $Q^*$  can be produced at minimal cost, *given* that  $W = \$10$  and  $C = \$20$ . No lower isoexpenditure curve touches the isoquant, meaning that  $Q^*$  cannot be produced for less than \$1,500.

An important characteristic of point  $Z$  is that the slope of the isoquant at point  $Z$  and the slope of the isoexpenditure line are the same (the slope of a curve at a given point is the slope of a line tangent to the curve at that point). The slope

of the isoquant at any given point is the *marginal rate of technical substitution* as defined in equation (3A.1). Another way of expressing equation (3A.1) is

$$MRTS = \frac{-\Delta K / \Delta Q}{\Delta L / \Delta Q} \quad (3A.2)$$

Equation (3A.2) directly indicates that the *MRTS* is a ratio reflecting the reduction of capital required to *decrease* output by one unit if enough extra labor is hired so that output is tending to *increase* by one unit. (The  $\Delta Q$ s in equation (3A.2) cancel each other and keep output constant.) Pursuing equation (3A.2) one step further, the numerator and denominator can be rearranged to obtain the following:<sup>2</sup>

$$MRTS = \frac{-\Delta K / \Delta Q}{\Delta L / \Delta Q} = -\frac{-\Delta Q / \Delta L}{\Delta Q / \Delta K} = -\frac{MP_L}{MP_K} \quad (3A.3)$$

where  $MP_L$  and  $MP_K$  are the marginal productivities of labor and capital, respectively.

The slope of the *isoexpenditure line* is equal to the negative of the ratio  $W/C$  (in Figure 3A.3,  $W/C$  equals  $10/20$ , or  $0.5$ ).<sup>3</sup> Thus, at point  $Z$ , where  $Q^*$  is produced in the minimum-cost fashion, the following equality holds:

$$MRTS = -\frac{MP_L}{MP_K} = -\frac{W}{C} \quad (3A.4)$$

Equation (3A.4) is simply a rearranged version of equation (3.8c) in the text.<sup>4</sup>

The economic meaning, or logic, behind the characteristics of cost minimization can most easily be seen by stating the *MRTS* as  $-\frac{\Delta K / \Delta Q}{\Delta L / \Delta Q}$  (see equation 3A.2) and equating this version of the *MRTS* to  $-\frac{W}{C}$ :

$$-\frac{\Delta K / \Delta Q}{\Delta L / \Delta Q} = -\frac{W}{C} \quad (3A.5)$$

or

$$\frac{\Delta K}{\Delta Q} \cdot C = \frac{\Delta L}{\Delta Q} \cdot W \quad (3A.6)$$

<sup>2</sup>This is done by making use of the fact that dividing one number by a second one is equivalent to multiplying the first by the *inverse* of the second.

<sup>3</sup>Note that  $10/20 = 75/150$ , or  $OB/OB'$ .

<sup>4</sup>The negative signs on each side of equation (3A.4) cancel each other and can therefore be ignored.

Equation (3A.6) makes it plain that to be minimizing costs, the cost of producing an extra unit of output by adding only labor must equal the cost of producing that extra unit by employing only additional capital. If these costs differed, the company could reduce total costs by expanding its use of the factor with which output can be increased more cheaply and cutting back on its use of the other factor. Any point where costs can still be reduced while  $Q$  is held constant is obviously not a point of cost minimization.

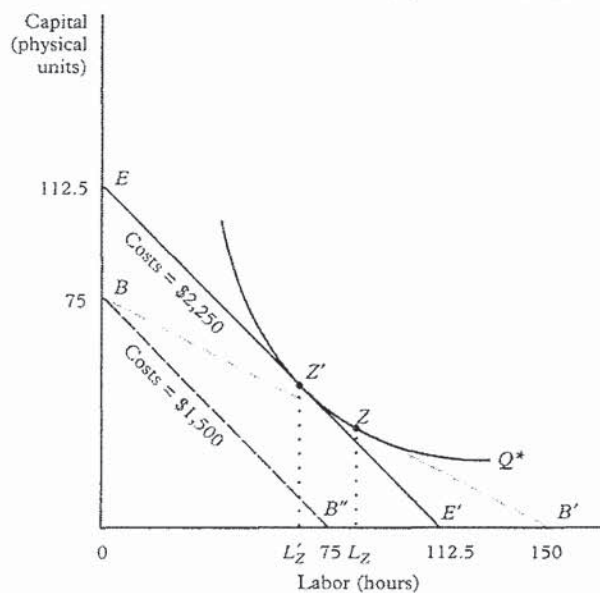
### The Substitution Effect

If the wage rate, which was assumed to be \$10 per hour in Figure 3A.3, goes up to \$20 per hour (holding  $C$  constant), what will happen to the cost-minimizing way of producing output of  $Q^*$ ? Figure 3A.4 illustrates the answer that common sense would suggest: total costs rise, and more capital and less labor are used to produce  $Q^*$ . At  $W = \$20$ , 150 units of labor can no longer be purchased if total costs are to be held to \$1,500; in fact, if costs are to equal \$1,500, only 75 units of labor can be hired. Thus, the isoexpenditure curve for \$1,500 in costs shifts from  $BB'$  to  $BB''$  and no longer is tangent to isoquant  $Q^*$ .  $Q^*$  can no longer be produced for \$1,500, and the cost of producing  $Q^*$  will rise. In Figure 3A.4, we assume the least-cost expenditure rises to \$2,250 (isoexpenditure line  $EE'$  is the one tangent to isoquant  $Q^*$ ).

Moreover, the increase in the cost of labor relative to capital induces the firm to use more capital and less labor. Graphically, the old tangency point of  $Z$  is replaced by a new one ( $Z'$ ), where the marginal productivity of labor is higher relative to  $MP_K$ , as our discussions of equations (3.8c) and (3A.4) explained. Point  $Z'$  is reached (from  $Z$ ) by adding more capital and reducing employment of labor. The movement from  $L_Z$  to  $L_{Z'}$  is the *substitution effect* generated by the wage increase.

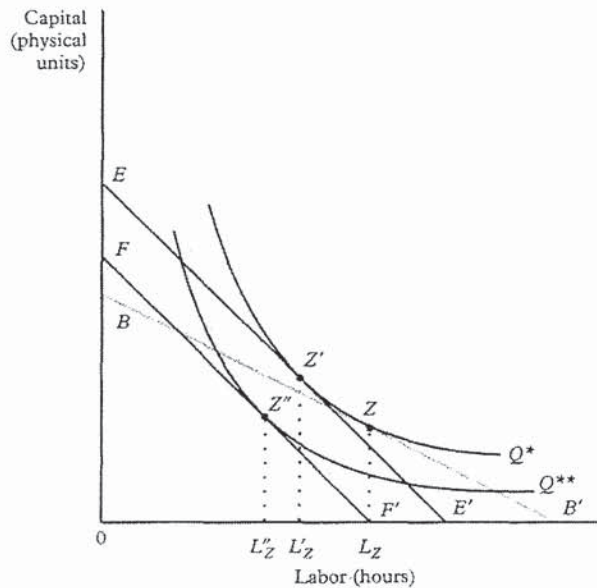
**FIGURE 3A.4**

Cost Minimization in the Production of  $Q^*$   
(Wage = \$20 per Hour; Price of a Unit of  
Capital = \$20)



**FIGURE 3A.5**

The Substitution and Scale Effects of a Wage Increase



### The Scale Effect

The fact that  $Q^*$  can no longer be produced for \$1,500, but instead involves at least \$2,250 in costs, will generally mean that it is no longer the profit-maximizing level of production. The new profit-maximizing level of production will be less than  $Q^*$  (how much less cannot be determined unless we know something about the product demand curve).

Suppose that the profit-maximizing level of output falls from  $Q^*$  to  $Q^{**}$ , as shown in Figure 3A.5. Since all isoexpenditure lines have the new slope of  $-1$  when  $W = \$20$  and  $C = \$20$ , the cost-minimizing way to produce  $Q^{**}$  will lie on an isoexpenditure line parallel to  $EE'$ . We find this cost-minimizing way to produce  $Q^{**}$  at point  $Z''$ , where an isoexpenditure line ( $FF'$ ) is tangent to the  $Q^{**}$  isoquant.

The *overall* response in the employment of labor to an increase in the wage rate has been a fall in labor usage from  $L_Z$  to  $L_Z''$ . The decline from  $L_Z$  to  $L_Z'$  is called the *substitution effect*, as we have noted. It results because the *proportions* of  $K$  and  $L$  used in production change when the ratio of wages to capital prices ( $W/C$ ) changes. The *scale effect* can be seen as the reduction in employment from  $L_Z'$  to  $L_Z''$ , wherein the usage of both  $K$  and  $L$  is cut back solely because of the reduced *scale* of production. Both effects are simultaneously present when wages increase and capital prices remain constant, but as Figure 3A.5 emphasizes, the effects are conceptually distinct and occur for different reasons. Together, these effects lead us to assert that the long-run labor demand curve slopes downward.